












⚡ This is an archived site and is no longer being updated. New 538 articles can be found at www.abcnews.com/538. >

APR. 3, 2008, AT 8:07 AM

Polls Now Weighted by Sample Size

By Nate Silver

Let me show you a catchy new graphic and then I'll explain what's going on.

Agency	Weight	Date	Size	Obama	Clinton
PA Quinnipiac	 2.10	3/28	3484	Obama +4	Clinton +8
PA Survey USA	 1.16	2/27	609	McCain +5	Clinton +1
PA Strat Vision	 0.96	3/10	1200	McCain +3	McCain +6
PA Rasmussen	 0.85	3/10	500	McCain +1	McCain +2
PA Susquehanna	 0.64	3/8	700	McCain +4	Clinton +3
PA Mansfield U.	 0.57	2/24	800	Obama +0.7	McCain +3.1
PA Quinnipiac	 0.54	2/23	1872	Obama +2	Clinton +2
PA Fr. Marshall	 0.46	2/18	640	McCain +1	Tie
PA Morning Call	 0.44	2/17	588	McCain +3	Clinton +1
PA Rasmussen	 0.30	2/14	500	Obama +10	McCain +2
PA Quinnipiac	 0.15	2/9	1419	Obama +1	Clinton +6
PA 538 Estimate				Obama +1.3	McCain +0.8
PA Average				Obama +0.1	Clinton +1.3
PA Win %				50.6%	56.3%

You'll see that, in addition to listing the date of the poll, we also list the sample size. And not only do we list the sample size, but we also weight the poll in part based on its sample size.

This is something I'd thought about doing all along, but a couple of the commenters had been prodding me on this issue in light of some recent polling that had particularly large sample sizes. The recent PPIC poll in California, for instance, had more than 1,000 respondents, and Quinnipiac's sample sizes have been enormous — they surveyed 3,484 people in their latest Pennsylvania poll!

Now, I don't quite assume a linear relationship between sample size and a poll's reliability. For one thing, different pollsters have different habits about how many respondents they include — all of Rasmussen's trial heat polls have included exactly 500 people, for example, while Strategic Vision favors sample sizes of either 800 or 1,200. And so, the information about the sample sizes may already be embedded to a certain extent in the pollster's [reliability rating](#), meaning that we might wind up double-counting the impact of sample size if we're not careful. For another thing, sample sizes are just one source of error, with methodological error perhaps be the

more important kind. I'm convinced that Zogby or ARG could survey all 325 million American citizens and still manage to fuck things up somehow.

Clearly though, if one poll is going to survey 3,484 voters and another poll is going to survey 348, the bigger survey ought to get some kind of extra credit. So the formula I use is

$$(\text{sample}/600)^{0.5}$$

e.g. the poll's sample size, divided by 600, and taken to its square root. The number 600 is chosen because it represents an "average" sample size; it also represents a margin of error of exactly +/- 4. And taking things to the square root has the effect of accounting for diminishing returns. The Quinnipiac poll of Pennsylvania comes out to 2.41 under this formula, for instance, whereas a typical Rasmussen poll of 500 respondents works out to 0.91.

So we now have **three** factors we use to determine the weight a poll is given in our model:

1. Its sample size, subject to the calculation I described above.
2. Its [pollster reliability rating](#).
3. Its recentness rating, as described in the [FAQ](#). Note that "old" polls from the same polling firm are punished under our recentness formula (FYI, I *have* validated based on data from previous election cycles that this tends to improve the accuracy of our estimates more than either throwing out "old" polls from the same agency, or not penalizing them at all).

The Quinnipiac poll, for instance, has a sample size rating of 2.41, a reliability rating of 0.95, and a recentness rating of 0.91; we multiply these three numbers together to get its overall weighting, which is 2.10. In the interests of transparency, I'm now explicitly listing the weighting I'm using for each and every poll in our data table (and I'm now calling it 'weight' rather than 'reliability', since the pollster's reliability score is just one of three factors we use in the weighting calculation).

Yes, I know this stuff gets complicated. But that's why you come here, right?

And it does turn out to make some difference. Texas, for instance, had been showing up on Obama's Swing State List in part because of a [Survey USA poll](#) that showed him trailing McCain by just 1 in that state. This is the most recent Survey USA poll, and Survey USA is a very reliable polling agency, so this poll does deserve some weight. However, this poll had a sample size of exactly 600, whereas there were a number of Texas polls, taken right around the same time, that have exceptionally large sample sizes. An IVR poll that showed Obama trailing McCain by 22 had more than 2,900

respondents; CNN conducted a poll with 1,500 respondents, and a previous Survey USA poll had 1,725. These large sample size polls now have a comparatively larger weight, and as such Obama's numbers have dropped a couple points in Texas, enough to take it off his Swing State List.

Overall, however, including the sample sizes has turned out to be helpful to both Democrats, particularly on the strength of the large-sample Quinnipiac polls in places like Ohio, Pennsylvania, and New Jersey.